引用格式:马秀盼,鲍娜,白涛,等. 基于多尺度双注意力的circRNA-miRNA相互作用预测方法[J]. 延安大学学报(自然科学版),2026,45(3):120-127. [MA X P, BAO N, BAI T, et al. Interaction prediction method of circRNA-miRNA based on multi-scale dual attention[J]. Journal of Yan'an University(Natural Science Edition),2026,45(3):120-127.] **DOI**:10.13876/J.cnki.ydnse.250111

基于多尺度双注意力的circRNA-miRNA相互作用预测方法

马秀盼,鲍　娜,白　涛*,王文东*

(延安大学 数学与计算机科学学院,陕西 延安 716000)

摘　要:环状RNA(circRNA)和微小RNA(miRNA)是两类重要的非编码RNA,在基因调控和疾病发病机制中发挥着重要作用,研究二者之间的相互作用对理解分子调控机制以及识别潜在的生物标志物具有重要意义。然而,现有的计算方法仍面临数据稀疏和多源特征异质性等问题,多局限于单一尺度建模和简单特征拼接,不足以捕捉特征间的交互与多尺度依赖。为解决上述问题,本研究提出了一种基于多尺度双注意力的多源特征融合方法MDACMI,该方法利用多尺度多头自注意力机制捕捉特征内部的局部和全局依赖关系,同时采用多尺度交叉注意力建立不同特征源之间的交互依赖,从而实现circRNA与miRNA在序列、属性和行为特征层面的深度融合。实验结果表明,MDACMI在3个公开数据集上的准确率分别达到0.917 7、0.962 8和0.957 8,AUC与AUPR值均超过0.97。进一步的t-SNE特征分析表明,该方法在复杂生物网络预测中具有较高的可靠性和实用价值。

关键词:circRNA-miRNA相互作用;交叉注意力机制;多头自注意力;多源特征表征融合

中图分类号:Q811.4　TP183　　**文献标志码**:A　　**文章编号**:1004-602X(2026)03-0120-08

随着生物信息学和测序技术的持续发展,研究人员在真核生物细胞中发现了大量功能多样的非编码RNA(non-coding RNA, ncRNA),这类RNA不直接参与蛋白质编码,却在基因调控中有着不可忽略的作用。在众多ncRNA中,环状RNA(circRNA)因具有独特的闭环结构脱颖而出,这种结构让它更加稳定,不易被核苷酸外切酶降解。正是这种稳定性,让circRNA在基因调控表达与疾病研究中扮演着重要角色[1]。研究表明,circRNA可以作为miRNA的海绵,通过结合微小RNA(miRNA)形成竞争性内源RNA网络来调控下游基因表达[2],进而影响细胞的生理和病理过程。这意味着,准确识别circRNA-miRNA相互作用(CMI)对于研究复杂疾病的产生、诊断和治疗具有十分重要的意义。

虽然传统生物湿实验方法能够有效验证CMI,但这类方法费时费力且成本高,难以对由数万样本组成的大规模样本逐一验证。因此,怎样利用计算方法来预测CMI成为近年来该领域的重要研究内容之一。目前主流的计算方法可以分为三类:第一类是以网络拓扑与相似性为预测思路。例如,NECMA[3]通过高斯相互作用谱核函数和NetMF算法学习circRNA、miRNA的潜在表示,并结合矩阵分解来预测CMI;CMIVGSD[4]使用奇异值分解与变分图自动编码器来融合线性与非线性特征,再由LightGBM进行分类。虽然这类方法可以利用已知网络拓扑结构提高拟合性能,但它们过度依赖已知的相互作用数据,难以挖

收稿日期:2025-10-09
基金项目:国家自然科学基金项目(62462062)
作者简介:马秀盼(2000—),女,硕士研究生,主要从事生物信息研究。
*通信作者　E-mail:yadxwwd@163.com

掘RNA序列中的原始特征信息。第二类方法开始引入序列信息。一个典型的方法是WANG等[5]提出的KGDCMI模型，该方法将K-mer序列特征、相似性特征和行为特征整合构建多源特征体系，并通过深度神经网络对CMI进行预测。与基于拓扑结构的方法相比，这一策略在一定程度上丰富了特征表达，但仍存在一定的局限性。第三类侧重于多源融合与更深层次语义的提取。例如，BCMCMI[6]先利用BERT从序列中提取语义特征，再通过Metapath2vec捕获高阶拓扑路径；BGF-CMAP[7]则采用Word2vec对序列进行嵌入表示，并与LINE及图分解相结合表征网络结构。这类方法在特征融合方面迈出一大步，但融合策略仍局限于简单拼接或传统分类器，而不能充分挖掘特征间的深层交互。并且大多数现有算法停留在单一尺度建模，这限制了模型捕捉局部细粒度信息和特征间全局依赖的能力。

考虑到上述方法存在的局限性，本文提出了一种基于多尺度双注意力的circRNA-miRNA相互作用预测方法——MDACMI。该方法结合序列内在信息、理化性质、属性特征及行为特征构建多源输入框架，再引入多尺度双注意力机制对特征进行深度融合。其中多尺度多头自注意力（MSSA）用来分别刻画不同尺度下的特征内的局部与全局依赖关系；多尺度交叉注意力（MSCA）则用来增强不同特征之间的信息交互，二者的结合使用有助于实现多层次信息的深度整合与判别。在多个数据集上的实验结果表明，MDACMI具有良好的预测性能与泛化能力。本文的主要贡献如下：

1）将序列内在信息与理化属性特征引入，与属性特征、行为特征构成多源特征输入框架，使得在预测CMI时的信息表征更全面。

2）设计了多尺度双注意力机制，目的在于挖掘特征内部依赖的同时捕获跨特征信息交互，使特征表达更具层次性与互补性。

3）在3个数据集上对MDACMI的性能进行了系统测试，验证其在不同数据场景下的稳定性与泛化能力。

1　数据与方法

1.1　数据

本研究采用3个广泛用于CMI预测的标准数据集，均来源于公共数据库CircR2Cancer[8]与CircBank[9]。其中，CircR2Cancer收录了经实验验证的circRNA-miRNA相互作用对，支持正负样本的构建；CircBank提供了circRNA的标准化命名、序列信息及潜在miRNA结合位点预测。基于上述数据库选择了3个不同规模的数据集：1）CMI-20208数据集：3 569个circRNA、1 152个miRNA和20 208条相互作用；2）CMI-9589数据集：2 115个circRNA、821个miRNA和9 589条相互作用；3）CMI-9905数据集：2 346个circRNA、962个miRNA和9 905条相互作用。

1.2　方法概述

MDACMI的框架图如图1所示，由特征输入、多尺度多头自注意力、多尺度交叉注意以及特征融合预测模块四部分组成。首先将提取的属性、行为和序列特征分别输入到多尺度多头自注意力机制，计算3个通道尺度下的单一特征的并行多头自注意力分布，再融合多尺度结果得到每类特征的多尺度表征。其中，细粒度尺度用来捕获细节信息和局部依赖关系，中等尺度则为了增强对局部特征间依赖的挖掘能力，而全局尺度旨在提取长程依赖和整体语义信息。之后再将三类特征的多尺度表征输入到多尺度交叉注意力，来捕获两两特征之间的交互，每个尺度下的属性、行为与序列特征分别生成自身的$\boldsymbol{Q}/\boldsymbol{K}/\boldsymbol{V}$，计算两两特征间的注意力分数，以挖掘3种不同特征间的依赖信息。接着利用加法归一化将MSSA输出的特征与MSCA生成的跨特征表示进行深度融合。最后输入到前馈网络进行非线性变换，由分类层对CMI进行最终预测。

1.3　特征提取

本研究采用多视角特征深度融合策略，分别提取序列、属性与行为三类特征对circRNA-miRNA相互作用进行表征。具体而言，将RNA序列经过work2vec编码生成语义嵌入向量，作为属性特征输入模型；根据已知的circRNA-miRNA相互作用关系对构建同质图，再通过图嵌入获得节点表示，作为行为特征输入；同时，提取circRNA与miRNA原始序列固有的内在信息与理化性质作为补充的序列特征。

1.3.1　序列特征提取

本文在从circRNA和miRNA的原始核苷酸序列中提取多维度的特征信息时，参考使用了文献[10]构建且公开提供的RNA序列特征提取代码。这组信息涵盖编码潜能、组成模式以及理化性质等多个方面，分别为开放阅读框长度、开放阅读框覆盖度、

六元组得分、Fickett得分、组成-转变-分布、等电点以及电子-离子相互作用势。共同构成了一组刻画circRNA和miRNA序列内在生物学特性的48维特征向量。

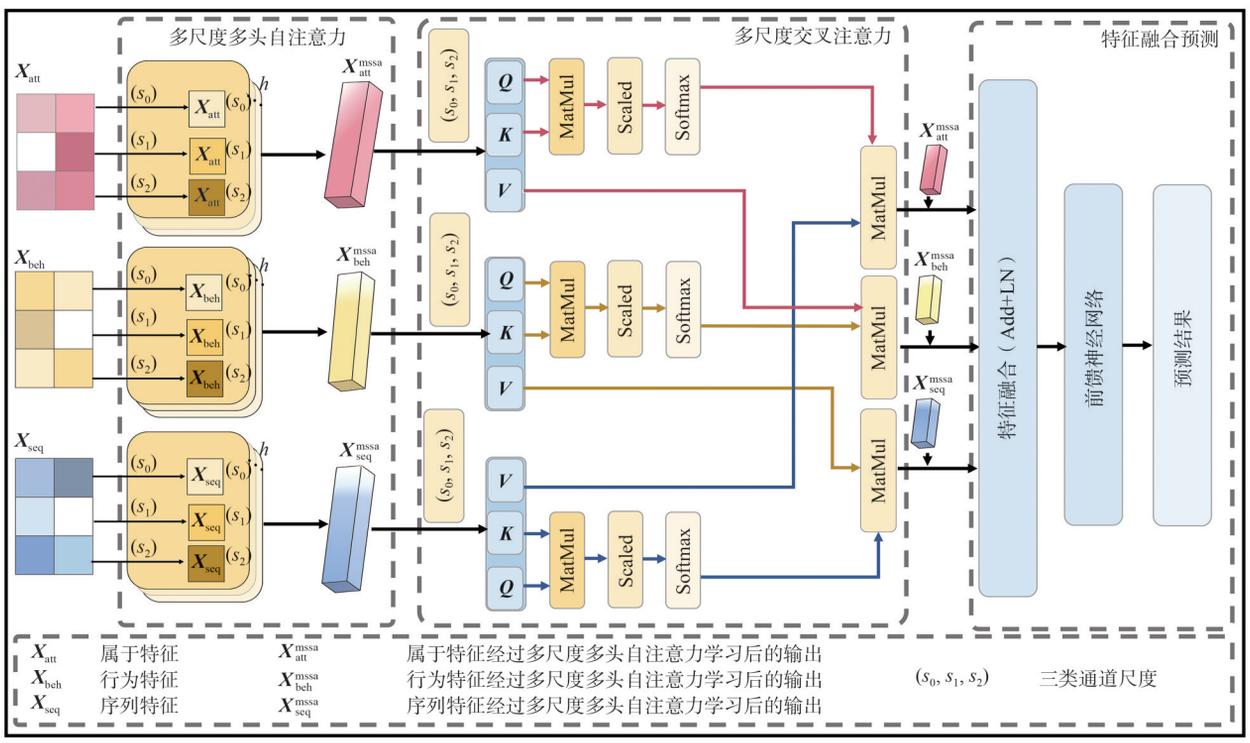


图1 MDACMI模型架构

1.3.2 属性特征提取

本文借鉴自然语言处理中的Word2Vec算法，将每条circRNA或miRNA核苷酸序列作为一个“句子”，将序列中的每个碱基(A、C、G、U)作为单独的“词”进行嵌入表示[7]。通过采用连续词袋模型，利用上下文碱基预测目标碱基，进而学习序列中的潜在语义依赖关系。其目标函数可表示为

$$\mathcal{L}_{\mathrm{CBOW}}=\frac{1}{N}\sum_{i=1}^{N}\log P\left(w_i|C_i\right), \tag{1}$$

其中，w_i为当前待预测的目标碱基，$C_i=\{w_{i-c},\cdots,w_{i-1},w_{i+1},\cdots,w_{i+c}\}$表示为目标碱基$w_i$上下文集合，$c$为上下文窗口大小，$N$为序列中碱基总数。

训练过程词向量维度设为64，其余参数均采用默认配置。最终得到64维嵌入向量用于表征序列的属性特征。

1.3.3 行为特征提取

行为特征由两种同构图嵌入方法获得。第一种是图分解方法，通过将节点间的连接强度映射为对应嵌入向量的内积来保留低维空间的网络局部结构关系[11]，目的在于引入正则化约束的同时最小化重构误差；第二种方法是大规模信息网络嵌入[12]，通过对数似然来优化节点嵌入的相似性。

通过上述两种特征处理方法，将每个circRNA与miRNA节点嵌入到统一的64维空间中，作为行为特征表示。

1.4 多尺度双注意力机制

1.4.1 多尺度多头自注意力

本文对传统多头自注意力[13]进行多尺度映射，利用多尺度多头自注意力模块来捕获特征的局部细节与全局依赖，以提升三类特征的判别能力。先对输入的每种特征进行不同尺度的线性变换，分别得到原始尺度(s_0)、中等尺度(s_1)和压缩尺度(s_2)的表示：

$$\boldsymbol{X}_i^{(s_0)}=\boldsymbol{X}_i, \tag{2}$$

$$\boldsymbol{X}_i^{(s_1)}=f\left(\boldsymbol{W}_1\boldsymbol{X}_i\right), \tag{3}$$

$$\boldsymbol{X}_i^{(s_2)}=f\left(\boldsymbol{W}_2\boldsymbol{X}_i\right), \tag{4}$$

其中，$\boldsymbol{X}_i\in\mathbb{R}^{n_i\times d}$，$i\in\{\mathrm{att},\mathrm{beh},\mathrm{seq}\}$是输入的第$i$类特征，$\boldsymbol{W}_1$和$\boldsymbol{W}_2$是不同尺度下可学习的投影矩阵，$f(\cdot)$是非线性激活函数。

每个尺度的特征表示会经过线性映射得到查询($\boldsymbol{Q}$)、键($\boldsymbol{K}$)和值($\boldsymbol{V}$)矩阵，再计算注意力分布[14-15]：

$$\boldsymbol{Q}^{(s_j)}=\boldsymbol{X}^{s}\boldsymbol{W}^{Q}\in\mathbb{R}^{n\times d_k}, \tag{5}$$

$$\boldsymbol{K}^{(s_j)}=\boldsymbol{X}^{s}\boldsymbol{W}^{K}\in\mathbb{R}^{n\times d_k}, \tag{6}$$

$$\boldsymbol{V}^{(s_j)} = \boldsymbol{X}^{s_j}\boldsymbol{W}^{V} \in \mathbb{R}^{n \times d_v}, \tag{7}$$

$$A(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\boldsymbol{K}}}}\right)\boldsymbol{V}, \tag{8}$$

其中，$d_{\boldsymbol{K}}$和$d_{\boldsymbol{V}}$分别代表键、值矩阵的维度，$A(\cdot)$表示注意力函数。

采用多头自注意力机制计算拼接后的最终多头自注意力输出表示：

$$Mul(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Concat(h_1, \cdots, h_h)\boldsymbol{W}^{O}, \tag{9}$$

其中，$h_i = A(\boldsymbol{Q}\boldsymbol{W}_i^{Q}, \boldsymbol{K}\boldsymbol{W}_i^{K}, \boldsymbol{V}\boldsymbol{W}_i^{V}), i = 1, 2, \cdots, h,$ (10)

$\boldsymbol{W}_i^{Q}, \boldsymbol{W}_i^{K}, \boldsymbol{W}_i^{V}$为第$i$个注意力头对应的线性投影矩阵，$\boldsymbol{W}^{O}$为输出投影矩阵。

最后将这些不同尺度下得到的注意力输出进行加权求和融合：

$$\boldsymbol{Z} = \sum_{j=0}^{2}\boldsymbol{U}^{j} \cdot Mul(\boldsymbol{Q}^{(s_j)}, \boldsymbol{K}^{(s_j)}, \boldsymbol{V}^{(s_j)}), \tag{11}$$

其中，$\boldsymbol{U}^{j}$是学习得到的映射矩阵。

1.4.2　多尺度交叉注意力

同样将多尺度映射到交叉注意力机制中，目的是建立两两特征间深层依赖的同时，捕获跨特征交互的细粒度局部信息与全局语义联系。与1.4.1类似，首先通过不同尺度的线性映射，得到原始尺度s_0、中等尺度s_1和压缩尺度s_2的表示[16]：

$$\boldsymbol{X}_i^{(s_0)} = \boldsymbol{X}_i, \boldsymbol{X}_i^{(s_1)} = f(\boldsymbol{W}_1\boldsymbol{X}_i), \boldsymbol{X}_i^{(s_2)} = f(\boldsymbol{W}_2\boldsymbol{X}_i), \tag{12}$$

对每个尺度s_j下的特征表示进行线性映射得到查询($\boldsymbol{Q}$)、键($\boldsymbol{K}$)和值($\boldsymbol{V}$)矩阵：

$$\boldsymbol{Q}^{(s_j)} = \boldsymbol{X}^{(s_j)}\boldsymbol{W}_i^{Q}, \tag{13}$$

$$\boldsymbol{K}^{(s_j)} = \boldsymbol{X}^{(s_j)}\boldsymbol{W}_i^{K}, \tag{14}$$

$$\boldsymbol{V}^{(s_j)} = \boldsymbol{X}^{(s_j)}\boldsymbol{W}_i^{V}, \tag{15}$$

其中，$\boldsymbol{W}_i^{Q}$、$\boldsymbol{W}_i^{K}$、$\boldsymbol{W}_i^{V}$分别为可学习的注意力权重。

以属性特征$\boldsymbol{X}_{\mathrm{att}}$为源模态、行为特征$\boldsymbol{X}_{\mathrm{beh}}$为目标模态计算跨特征注意力为例，计算特征间的交叉注意力分布：

$$CA^{(s_j)}(\mathrm{att} \leftarrow \mathrm{beh}) = softmax\left(\frac{\boldsymbol{Q}_{\mathrm{att}}^{(s_j)}(\boldsymbol{K}_{\mathrm{beh}}^{(s_j)})^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\mathrm{beh}}^{s_j}, \tag{16}$$

其中，d_k是用于缩放点积以保证数值稳定性的键向量维度。

最后，把各尺度下的交叉注意力结果投影回原始维度进行融合，具体计算公式如下：

$$\boldsymbol{Z}_{\mathrm{att} \rightarrow \mathrm{beh}} = \sum_{j=0}^{2}\boldsymbol{U}^{(s_j)} \cdot CA^{(s_j)}(\mathrm{att} \leftarrow \mathrm{beh}), \tag{17}$$

其中，$\boldsymbol{U}^{(s_j)}$是融合权重矩阵。

1.4.3　多尺度双注意力融合

本文设计双注意力驱动的融合策略，以整合两种机制的互补信息：

$$\boldsymbol{X}_i = LayerN(\boldsymbol{X}_i^{\mathrm{mssa}} + \boldsymbol{X}_i^{\mathrm{cross}}), i \in \{\mathrm{att}, \mathrm{beh}, \mathrm{seq}\}, \tag{18}$$

其中，$\boldsymbol{X}_i^{\mathrm{mssa}}$为第$i$种特征的MSSA输出表示，$\boldsymbol{X}_i^{\mathrm{cross}}$为第$i$种特征通过MSCA的输出表示。

引入前馈神经网络，进一步提升融合表示的非线性建模能力：

$$FFN(\boldsymbol{X}_i) = Relu(\boldsymbol{X}_i\boldsymbol{W}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2, \tag{19}$$

其中，$\boldsymbol{X}_i$为融合特征的第i个输出特征，$Relu(\cdot)$为非线性激活函数，$\boldsymbol{W}_1$、$\boldsymbol{b}_1$和$\boldsymbol{W}_2$、$\boldsymbol{b}_2$分别为第一、第二层的权重与偏置。

1.5　相互作用预测

将已经完成多源特征深度融合的三类特征拼接为统一向量：

$$\boldsymbol{F}_{\mathrm{fused}} = [\boldsymbol{F}_{\mathrm{seq}} \| \boldsymbol{F}_{\mathrm{att}} \| \boldsymbol{F}_{\mathrm{beh}}]。 \tag{20}$$

通过全连接层对拼接后的向量进行非线性变换，增强特征的表达能力：

$$h = \sigma(\boldsymbol{W}_h\boldsymbol{F}_{\mathrm{fused}} + \boldsymbol{b}_h), \tag{21}$$

其中，$\|$表示特征拼接操作，$\boldsymbol{W}_h$和$\boldsymbol{b}_h$分别为可学习的权重参数与偏置参数，σ是非线性激活函数。

最后输入到输出层，得到CMI预测概率：

$$\hat{y} = \frac{1}{1 + \mathrm{e}^{-(\boldsymbol{W}_o\boldsymbol{F}_{\mathrm{fused}} + b_o)}}, \tag{22}$$

其中，$\hat{y} \in [0, 1]$表示预测的CMI概率，$\boldsymbol{W}_o$和b_o分别是输出层的权重矩阵与偏置项。

1.6　评价指标

本文选用了7个常用评价指标对MDACMI进行评估，分别为准确率(Accuracy，ACC)δ_{ACC}、灵敏度(Sensitivity，Sen)δ_{Sen}、特异性(Specificity，Spe)δ_{Spe}、精确率(Precision，Prec)δ_{Prec}、马修斯相关系数(Matthews Correlation Coefficient，MCC)δ_{MCC}、精确率-召回率曲线下面积(AUPR)δ_{AUPR}以及受试者工作特征曲线下面积(AUC)δ_{AUC}。详细的计算公式如下：

$$\delta_{\mathrm{ACC}} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{23}$$

$$\delta_{\mathrm{Sen}} = \frac{TP}{TP + FN}, \tag{24}$$

$$\delta_{\mathrm{Spe}} = \frac{TN}{TN + FP}, \tag{25}$$

$$\delta_{\mathrm{Prec}} = \frac{TP}{TP + FP}, \tag{26}$$

$$\delta_{MCC}=\frac{TP\cdot TN-FP\cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}},\tag{27}$$

其中,TP、TN、FP和FN分别表示真阳性、真阴性、假阳性和假阴性的数量。

2 实验对比与分析

2.1 多头自注意力头数参数优化

下面对注意力的头数参数进行实验分析,以确定多头自注意力机制的最佳配置,在头数为2、4、6、8、10和12的条件下进行了实验对比(见表1)。由表1可知,头数较少时的ACC、MCC、AUPR等评价指标偏低,这可能因为特征被划分的子空间有限,导致模型难以充分捕获多层次信息;而当头数增加至10和12时的性能指标依然不佳,这可能由于过多的子空间使单个注意力头接收的信息量减少,增加了计算开销的同时还弱化了注意力的整体表达能力。所以,综合模型在不同头数下的整体性能表现,本研究最终选择8个注意力头作为模型的最优配置。

表1 多尺度多头注意力头数参数分析

头数	δ_{ACC}	δ_{Sen}	δ_{Spe}	δ_{Prec}	δ_{MCC}	δ_{AUPR}	δ_{AUC}
12	0.912 8	0.899 2	0.926 4	0.924 6	0.826 2	0.970 8	0.969 5
10	0.912 9	0.893 0	**0.932 8**	**0.930 2**	0.826 7	**0.971 7**	0.970 1
8	**0.917 7**	**0.915 0**	0.920 5	0.920 2	**0.835 7**	0.971 0	**0.972 2**
6	0.911 5	0.904 7	0.918 2	0.917 2	0.823 1	0.970 4	0.969 2
4	0.911 2	0.901 2	0.921 3	0.920 0	0.823 0	0.969 0	0.967 9
2	0.909 6	0.899 3	0.920 0	0.918 5	0.819 6	0.969 2	0.968 3

注:最佳性能值用粗体标出。

2.2 五折交叉验证性能评估

本研究采用五折交叉验证在3个数据集上对MDACMI进行系统评估,本文的所有实验结果均为五折结果的平均值,目的是为了减小数据划分带来的偶然性影响。

表2展示了在数据集CMI-20208上的五折实验结果,可以观察到其平均准确率达到0.917 7,AUC和AUPR分别为0.972 2和0.971 0,Spe、MCC等评价指标也保持着一致较高水平,表现了MDACMI良好的预测性能。在数据集CMI-9905和CMI-9589上进行的五折交叉验证结果,如知网增强出版附件表1和表2所示。知网增强出版附件图1展示了MDACMI在3个不同规模数据集上绘制的精确率-召回率曲线和受试者工作特征曲线。从图中可以看出本研究方法在三组数据集上都有着较低的标准差,且每一折之间的曲线走势高度一致,进一步证实了该方法在CMI预测任务中的稳定性与可重复性。

表2 MDACMI在数据集CMI-20208进行五折交叉验证的性能

折数	δ_{ACC}	δ_{Sen}	δ_{Spe}	δ_{Prec}	δ_{MCC}	δ_{AUPR}	δ_{AUC}
Fold1	0.916 6	0.928 3	0.905 0	0.907 2	0.833 5	0.972 0	0.971 7
Fold2	0.918 6	0.911 9	0.925 3	0.924 3	0.837 3	0.973 5	0.971 5
Fold3	0.917 1	0.914 9	0.919 3	0.919	0.834 2	0.969 9	0.968 3
Fold4	0.916 6	0.897	0.936 1	0.933 5	0.833 8	0.973 8	0.971 5
Fold5	0.920 0	0.923 3	0.916 8	0.917 4	0.840 1	0.971 7	0.972 0
Mean	0.917 7	0.915 0	0.920 5	0.920 2	0.835 7	0.971 0	0.972 2

2.3 与其他注意力机制对比

为了验证本文提出的双注意力机制的有效性,下面将MDACMI与4种注意力变体进行对比,其中包括通道注意力(SE、ECA)、空间注意力(CBAM)以及单尺度双注意力(MDACMI-SS)。在对比过程中,仅用各注意力替换原模型中的多尺度双注意力模块,其余结构通通保持一致。实验结果如表3所示,MDACMI在7个评价指标上的表现一致最佳。由表中实验结果可知,多尺度双注意力机制在不同尺度下更好的挖掘了特征内和特征间的依赖关系,增强了模型对局部信息与全局结构的关注度以及对细节和上下文信息的理解能力。

表3　不同注意力机制的比较

方法	δ_{ACC}	δ_{Sen}	δ_{Spe}	δ_{Prec}	δ_{MCC}	δ_{AUPR}	δ_{AUC}
MDACMI-SE	0.855 4	0.845 1	0.865 8	0.863 2	0.711 4	0.930 8	0.927 9
MDACMI-CBAM	0.850 3	0.836 1	0.864 6	0.860 8	0.701 3	0.927 5	0.924 4
MDACMI-ECA	0.836 4	0.802 4	0.870 4	0.861 3	0.674 7	0.917 7	0.911 8
MDACMI-SS	0.907 7	0.895 5	0.919 9	0.918 0	0.815 9	0.962 9	0.962 3
MDACMI	**0.917 7**	**0.915 0**	**0.920 5**	**0.920 2**	**0.835 7**	**0.971 0**	**0.972 2**

注：最佳性能值用粗体标出。

2.4　消融实验分析

在CMI-20208数据集上设计了消融实验，分别构建了3个模型变体与MDACMI本体进行对比，分析不同组件对特征表征能力的影响（见表4）。从表中可以观察到，直接去掉多尺度双注意力模块方法的整体性能明显下降；当只使用MSSA时，方法的平均准确率由0.883 4提升至0.911 3，相对提升约3.16%；而仅使用MSCA时，准确率为0.902 3，相比第一个变体增幅约2.14%；当同时引入MSSA与MSCA时的方法的平均准确率达到0.917 7，相比较第一个变体提升3.89%；AUC和AUPR分别从0.951 7和0.952 8提升至0.971 0和0.972 2。由此可见，同时引入MSSA与MSCA能够更好地捕捉circRNA与miRNA之间的复杂相互作用，二者的结合使用显著提升了MDACMI的整体预测性能，从而验证了该方法设计的合理性与有效性。

表4　MDACMI在数据集CMI-20208上模块消融评估

模型变体		δ_{ACC}	δ_{Sen}	δ_{Spe}	δ_{Prec}	δ_{MCC}	δ_{AUPR}	δ_{AUC}
MSCA	MSSA							
×	×	0.883 4	0.876 8	0.889 9	0.888 5	0.766 8	0.952 8	0.951 7
×	√	0.911 3	0.897 0	**0.925 6**	**0.923 5**	0.823 0	0.970 0	0.968 9
√	×	0.902 3	0.883 4	0.921 4	0.918 4	0.805 5	0.965 9	0.964 6
√	√	**0.917 7**	**0.915 0**	0.920 5	0.920 2	**0.835 7**	**0.972 2**	**0.971 0**

注：最佳性能值用粗体标出，√表示包含该模块，×表示不包含该模块。

2.5　特征可视化分析

下面从特征可视化角度展开验证MDACMI的有效性，实验采用t-SNE方法对原始特征及经过模型融合后的特征分布进行降维且展示，具体效果见知网增强出版附件图2。从可视化结果可以看到，经多尺度双注意力机制深度融合后的特征，在低维空间中的正负样本呈现出明显的聚簇分布，形成了清晰且稳定的类间分离结构。

2.6　与基准方法对比

为了全面评估本文方法在CMI预测任务中的性能，本文在完全一致的实验设置下，将MDACMI与以下14种基准方法进行了对比，这些方法涵盖了当前CMI预测领域中被广泛采用且具有代表性的算法：WSCD[17]、KGDCMI[5]、SGCCMI[18]、BGF-CMAP[7]、CA-CMA[19]、LSAGHI[20]、DGCLCMI[21]、MFERL[22]、JSNDCMI[23]、KS-CMI[24]、GCNCMI[25]、MuGNet-CMI[26]、iHofman[27]以及HCLAMCMI[28]。

本文在3个不同数据集上展开对比实验（见表5）。从实验结果可以看出，MDACMI在AUC和AUPR两个指标上一致优于已有的对比方法。在CMI-20208数据集上，与MFERL相比，AUC和AUPR分别提升0.025 4和0.027 9；在CMI-9589数据集上，相较于性能最好的HCLAMCMI，AUC提升0.019 4、AUPR提升0.015 6；在CMI-9905数据集上与HCLAMCMI相比，AUC和AUPR分别提升0.027 4、0.021 5。这一结果不仅证明了MDACMI在不同大小、不同数据分布的数据集上均能保持良好的预测性能与泛化能力，而且对该方法的有效性提供了有力支持。

表5　MDACMI与基准方法的性能对比

CMI-20208			CMI-9589			CMI-9905		
模型	δ_{AUPR}	δ_{AUC}	模型	δ_{AUPR}	δ_{AUC}	模型	δ_{AUPR}	δ_{AUC}
WSCD	0.884 7	0.889 8	KGDCMI	0.893 7	0.904 1	KGDCMI	0.876 7	0.893 0
KGDCMI	0.876 7	0.893 0	CACMA	0.908 6	0.915 6	SGCNCMI	0.888 7	0.894 2
SGCCMI	0.888 7	0.894 2	KS-CMI	0.918 1	0.917 9	JSNDCMI	0.899 9	0.903 3
BGF-CMAP	0.912 0	0.907 5	GCNCMI	0.939 6	0.932 0	KS-CMI	0.914 4	0.908 6
CA-CMA	0.913 1	0.917 0	JSNDCMI	0.940 3	0.941 5	CA-CMA	0.908 8	0.913 8

(续表5)

CMI-20208			CMI-9589			CMI-9905		
模型	δ_{AUPR}	δ_{AUC}	模型	δ_{AUPR}	δ_{AUC}	模型	δ_{AUPR}	δ_{AUC}
LSAGHI	0.947 9	0.952 2	LSAGHI	0.952 0	0.955 2	LSAGHI	0.943 5	0.940 5
DGCLCMI	0.941 5	0.954 6	DGCLCMI	0.950 1	0.961 0	DGCLCMI	0.954 6	0.964 5
MFERL	0.944 3	0.945 6	MFERL	0.962 9	0.966 9	MFERL	0.961 0	0.959 6
MuGNetCMI	0.918 7	0.922 1	MuGNetCMI	0.957 7	0.956 8	MuGNetCMI	0.927 8	0.924 8
iHofman	0.921 1	0.917 3	iHofman	0.926 9	0.918 7	iHofman	0.923 4	0.909 5
–	–	–	HCLAMCMI	0.974 9	0.971 7	HCLAMCMI	0.968 3	0.964 4
MDACMI	**0.972 2**	**0.971 0**	MDACMI	**0.990 5**	**0.991 1**	MDACMI	**0.989 8**	**0.991 8**

注:最佳性能值用粗体标出;对比方法的结果来源于各自原文献报道性能,均基于相同公开数据集。

3 结论与展望

针对CMI预测任务中多源异质特征融合不足以及跨特征依赖关系复杂的问题,本文提出了一种基于多尺度双注意力的多源特征融合的预测方法MDACMI。该方法从属性、行为与序列3种特征出发进行建模,利用多尺度自注意力机制对单一特征内部的细粒度与全局信息进行挖掘,随后通过多尺度交叉注意力机制实现不同特征间的语义交互与整合。通过两种注意力机制的配合,MDACMI能够同时兼顾局部特征表达与全局关联信息表征。实验结果表明,本文所提方法显著增强了模型对circRNA-miRNA相互作用的建模能力,证实了该模型的有效性。

虽然MDACMI在circRNA与miRNA相互作用预测任务中取得了显著进展,但仍存在进一步拓展的空间。未来的研究工作可以从以下两个方面进行:1)当前方法主要采用的是固定尺度投影,后续研究可以尝试自适应尺度选择策略或多层次动态融合来提高模型的灵活性;2)本文仅在CMI预测任务上对模型分类性能进行了验证,后续可将该方法扩展到其他生物分子相互作用预测的问题上验证模型的通用性,比如蛋白质-RNA、lncRNA-miRNA等。

特别提醒:本文的附件材料见知网增强出版附件。

参考文献:

[1]MEMCZAK S, JENS M, ELEFSINIOTI A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. Nature, 2013, 495(7441): 333-338.

[2]HANSEN T B, JENSEN T I, CLAUSEN B H, et al. Natural RNA circles function as efficient microRNA sponges [J]. Nature, 2013, 495(7441): 384-388.

[3]LAN W, ZHU M, CHEN Q, et al. Prediction of circRNA-miRNA associations based on network embedding [J]. Complexity, 2021, 2021(1): 6659695.

[4]QIAN Y, ZHENG J, ZHANG Z, et al. CMIVGSD: circRNA-miRNA interaction prediction based on variational graph auto-encoder and singular value decomposition[C]//2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Piscataway: IEEE, 2021: 1655-1662.

[5]WANG X F, YU C Q, LI L P, et al. KGDCMI: A new approach for predicting circRNA-miRNA interactions from multi-source information extraction and deep learning [J]. Frontiers in Genetics, 2022, 13: 958096.

[6]WEI M M, YU C Q, LI L P, et al. BCMCMI: A fusion model for predicting circRNA-miRNA interactions combining semantic and meta-path[J]. Journal of Chemical Information and Modeling, 2023, 63(16): 5384-5394.

[7]GUO L X, WANG L, YOU Z H, et al. Biolinguistic graph fusion model for circRNA-miRNA association prediction[J]. Briefings in Bioinformatics, 2024, 25(2): bbae058.

[8]LAN W, ZHU M, CHEN Q, et al. CircR2Cancer: a manually curated database of associations between circRNAs and cancers[J]. Database, 2020, 2020: baaa085.

[9]LIU M, WANG Q, SHEN J, et al. Circbank: a comprehensive database for circRNA with standard nomenclature[J]. RNA Biology, 2019, 16(7): 899-905.

[10]LIU T, ZOU B, HE M, et al. LncReader: identification of dual functional long noncoding RNAs using a multi-head self-attention mechanism[J]. Briefings in Bioinformatics, 2023, 24(1): bbac579.

[11]LI R, YUAN X, RADFAR M, et al. Graph signal processing, graph neural network and graph learning on biological data: a systematic review[J]. IEEE Reviews in Biomedical Engineering, 2021, 16: 109-135.

[12]GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: A survey[J]. Knowledge-Based Systems, 2018, 151: 78-94.

[13]ZHAO X, PENG X, NIU K, et al. A multi-head self-attention deep learning approach for detection and recommendation of neuromagnetic high frequency oscillations in epilepsy[J]. Frontiers in Neuroinformatics, 2022, 16: 771965.

[14]WANG H, HUANG T, WANG D, et al. MSCAN: multi-scale self- and cross-attention network for RNA methylation site prediction[J]. BMC Bioinformatics, 2024, 25(1): 32.

[15]WANG Y J, ZHAO Y P, JIN Y. Multi-level and multi-

scale cross attention network of wavelet packet transform for supersonic inlet unstart prediction[J]. Expert Systems with Applications,2025,272:126782.
[16]HUA W,ZHOU C,WU J,et al. MSVIT:Improving spiking vision transformer using multi-scale attention fusion[C]// Proceedings of the Thirty-fourth International Joint Conference on Artificial Intelligence (IJCAI-25). Montreal,2025:5399-5407.
[17]GUO L X,YOU Z H,WANG L,et al. A novel circRNA-miRNA association prediction model based on structural deep neural network embedding[J]. Briefings in Bioinformatics,2022,23(5):bbac391.
[18]YU C Q,WANG X F,LI L P,et al. SGCNCMI:A new model combining multi-modal information to predict circRNA-related miRNAs,diseases and genes[J]. Biology,2022,11(9):1350.
[19]GUO L X,WANG L,YOU Z H,et al. Likelihood-based feature representation learning combined with neighborhood information for predicting circRNA-miRNA associations[J]. Briefings in Bioinformatics,2024,25(2):bbae020.
[20]YIN W,WANG S,ZHANG Y,et al. Autonomously adjusting multi-relational hypergraphs structure for predicting circRNA-MiRNA associations[J]. IEEE Journal of Biomedical and Health Informatics,2025:1-10.
[21]CAO C,LI M,WANG C,et al. DGCLCMI:a deep graph collaboration learning method to predict circRNA-miRNA interactions[J]. BMC Biology,2025,23(1):104.
[22]PENG L,WANG W,YANG Z,et al. Leveraging explainable multi-scale features for fine-grained circRNA-miRNA interaction prediction[J]. BMC Biology,2025,23(1):121.
[23]WANG X F,YU C Q,YOU Z H,et al. A feature extraction method based on noise reduction for circRNA-miRNA interaction prediction combining multi-structure features in the association networks[J]. Briefings in Bioinformatics,2023,24(3):bbad111.
[24]WANG X F,YU C Q,YOU Z H,et al. KS-CMI:a circRNA-miRNA interaction prediction method based on the signed graph neural network and denoising autoencoder[J]. Iscience,2023,26(8):107478.
[25]HE J,XIAO P,CHEN C,et al. GCNCMI:A graph convolutional neural network approach for predicting circRNA-miRNA interactions[J]. Frontiers in Genetics,2022,13:959701.
[26]JIANG C,WANG L,YU C,et al. MuGNet-CMI:Multi-head hybrid graph neural network for predicting circRNA-miRNA interactions with global high-order and local low-order information[J]. IEEE Transactions on Big Data,2026,12(1):159-173.
[27]YU C Q,JIANG C,WANG L,et al. IHofman:A predictive model integrating high-order and low-order features with weighted attention mechanisms for circRNA-miRNA interactions[J]. BMC Biology,2025,23(1):162.
[28]CHEN L,CHEN Y,ZHOU B. HCLAMCMI:Prediction of circRNA-miRNA interactions based on hypergraph contrastive learning and an attention mechanism[J]. Journal of Chemical Information and Modeling,2025,65(21):12099-12115.

[责任编辑　毕　伟]

Interaction prediction method of circRNA-miRNA based on multi-scale dual attention

MA Xiupan,BAO Na,BAI Tao*,WANG Wendong*

(College of Mathematics and Computer Science, Yan'an University, Yan'an 716000, China)

Abstract:Circular RNA (circRNA) and microRNA (miRNA) are pivotal non-coding RNAs exerting vital roles in gene regulation and disease pathogenesis. Exploring their interactions is critical for elucidating molecular regulatory mechanisms and identifying potential biomarkers. However, existing computational methods suffer from data sparsity and multi-source feature heterogeneity, and are limited to single-scale modeling and simple feature concatenation, failing to capture feature interactions and multi-scale dependencies. To address these issues, this study proposes MDACMI, a multi-source feature fusion method based on multi-scale dual attention. Specifically, it uses multi-scale multi-head self-attention to capture intra-feature local and global dependencies, and multi-scale cross-attention to build cross-source feature interactions, achieving deep fusion of circRNA and miRNA sequence, attribute and behavioral features. Experiments on three public datasets show that MDACMI achieves accuracies of 0.917 7, 0.962 8 and 0.957 8, with AUC and AUPR both exceeding 0.97. Further t-SNE features analysis verifies that this method has high reliability and practical value in the prediction of complex biological network.

Key words:circRNA-miRNA interactions; multi-source feature fusion; cross-attention mechanism; multi-head self-attention